Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 1999 Proceedings

Americas Conference on Information Systems (AMCIS)

December 1999

Better Analysis by Analogy: Applying Data Modeling Techniques to Grounded Theory

D. Harrison McKnight
Florida State University

Gordon Everest
University of Minnesota

Gordon Davis
University of Minnesota

Better Analysis by Analogy: Applying Data Modeling Techniques to Grounded Theory

D. Harrison McKnight, Florida State University, dmcknigh@garnet.acns.fsu.edu
Gordon C. Everest and Gordon B. Davis, University of Minnesota, geverest/gdavis@csom.umn.edu

Abstract

When done well, grounded theory is an excellent qualitative research tool for explaining new or complex phenomena. Still, grounded theory is interpretative in nature, raising questions about the levels of reliability and validity in grounded theory studies. This paper argues that since grounded theory analysis methods are analogous to conceptual data modeling methods, data modeling techniques may be employed to enhance the validity and reliability of grounded theory research.

Data Modeling and Grounded Theory: Analogous Methods

Grounded theory (Glaser & Strauss, 1967) is a post-positivist research approach used to build and test social theory. Logical data modeling is the process used by system development professionals to represent the data requirements of an application system (March, 1992). Even though their purposes differ widely, grounded theory research and data modeling are similar processes in the following seven ways.

First, they both represent the meaning of abstract concepts (or constructs) as they relate to each other in the real world (Brown, 1993; Glaser & Strauss, 1967). Grounded theory represents meaning by modeling which concepts are causally related to other concepts in the context (Strauss & Corbin, 1990: 29). Data modeling represents meaning via "…things and their relationships" (Brown, 1993, p. 395).

Second, both are based on information gathered about the situation. In the analysis phase, data modelers gather information about organizational data needs to create entity types by analyzing how to classify the things about which data need to be stored. Grounded theory is also based on conceptual information (Strauss & Corbin, 1994: 273). Grounded theorists, like data modelers, try to discover, through interviews and observations, what exists in the domain of interest. They induce theoretical concepts from "open coding" of this information and then relate the concepts to each other using "axial coding." Information (qualitative data) drives the creation of the model.

Third, both grounded theory and data modeling use the abstract concept to represent populations of things. In grounded theory, these abstractions are typically called "categories" (Strauss & Corbin, 1990) or "concepts" (Strauss & Corbin, 1994). In data modeling, the abstractions are called "entity types." Both methods represent categories of things that exist as real world instances (March, 1992; Glaser & Strauss, 1967).

Fourth, both methods employ descriptors of the abstract concept. Data modeling uses attributes, which are characteristics of the entity (e.g., for a student entity: student name, ID number, address). Similarly, grounded theory specifies the properties of the category, such as "who…what…how… when…" (Strauss & Corbin, 1990: 79). Just as instances of data attributes may vary over some range of values (e.g., Julian dates: 1-366), so grounded theory properties are analyzed for dimensional range (Table 1).

Table 1 Grounded Theory Properties and their Range

Category	Properties	Dimensional Range (applied to each incident)
Watching	Frequency	Often ---------- Never
	Extent	More ---------- Less
	Intensity	High ---------- Low
	Duration	Long ---------- Short

Source: Strauss & Corbin, 1990: 72

Fifth, both methods abstract concepts to lower- and higher levels of abstraction. Data modeling uses supertypes and subtypes (e.g., employee--clerical employee), while grounded theory uses constructs and subconstructs (e.g., surgery outcome—relief). These are equivalent methods of representing hierarchies of related sets of real world things.

Sixth, both relate concepts to other concepts in a model, though the manner of relating differs. Grounded theory concepts relate via correlational or causal models (Strauss & Corbin, 1994). By contrast, data modeling entities relate to each other in terms of optionality (optional, mandatory), cardinality (e.g., 1:1; 1:M), and degree (e.g., unary, binary) (Hoffer, George & Valacich, 1996).

Seventh, both methods embody techniques for assuring the model is valid. Data modeling entity types represent "individual things. A thing has a boundary, an identity, and is distinct from all other things" (1993: 399). In data modeling, the analyst 'normalizes' an entity type to be sure that each of its non-key attributes "provide a fact about the key, the

whole key, and nothing but the key" (Kent, 1983: 121). This assures that each entity is valid: that it is both consistent internally and is distinct from all other entities—thus displaying convergent and discriminant validity.

Grounded theory uses a validity technique called the constant comparative method. Researchers ask, "to what class of phenomenon does [this instance] seem to pertain, and is it similar or different from [other classes]?" (Strauss & Corbin, 1990: 66). To determine where to place it, the instance is compared to the existing categories (Glaser & Strauss, 1967), and is compared with previous instances placed in the category. From these comparisons, theoretical properties of the category are generated. Researchers are to theorize about each construct instance, ensuring that each instance only belongs to one construct (Glaser & Strauss, 1967: 108). Then they integrate categories and their properties by comparing each instance to properties of the category. As with normalization, this assures that the category will represent a single, unitary thing.

Given the above analogies between grounded theory and data modeling, it is appropriate to view them as similar analytical processes. To the extent that they are similar, techniques from one method may benefit the other. We next explore ways in which data modeling techniques may profit grounded theory analyses.

Strengthening Grounded Theory Reliability and Construct Validity via Data Modeling Principles/Tools

Grounded theory has been praised for its rigor as a qualitative method: "...grounded theorists... are hard-nosed empiricists, system builders, ...and skeptical of nonsystematic theory and empirical work" (Denzin, 1994: 511). Still, Ives (1981) criticized grounded theorists regarding validity "… their basic answer is, they have experienced it…[they] assert validity, without defining how it is achieved or tested" (Ives, 1981: 360). Standards of analysis quality are currently undefined. Further, although grounded theory can be done in an exacting manner, additional rigor may improve model validity and reliability.

Construct validity means that the measures of a construct correspond closely to the conceptual meaning of the construct (Schwab, 1980), in terms of convergent and discriminant validity. Convergent validity means the extent to which responses from different measurements of the same construct are highly intra-correlated (Schwab, 1980). Discriminant validity means the extent to which a construct is distinct from other constructs. In Grounded theory research, reliability means coding repeatability.

Concerns about Grounded Theory. Although grounded theory, when done well (e.g., Orlikowski, 1993), is a valid qualitative research tool, it is interpretative in nature (Strauss & Corbin, 1990), and interpretation raises questions about the levels of reliability and validity of the resulting theoretical model. The amount of the data also raises concerns, because field notes and memos may become very voluminous and inflexible (Glaser, 1978; Ives, 1981). Hence, grounded theory methods pose (at least) five concerns: 1. a single data instance may inadvertently be interpreted as more than one construct; 2. readers can't tell how well the researchers handled the unwieldy volume of information; 3. grounded theory models may not be convincing because the reader cannot be anywhere near as familiar with the volumes of supporting data as is the researcher; 4. a second interpretation of a data segment may yield a different result than the first interpretation (reliability); and 5. no solid standards indicate precisely when a grounded theory model becomes adequately valid (in data modeling terms, fully 'normalized').

Suggestions to Address Concerns. To address these five concerns, the following suggestions are offered. Each suggestion applies data modeling techniques to grounded theory analysis.

Define each concept in a glossary, as recommended for data modeling entities (Brown, 1993: 292), addressing concern 1. A formal definition of the concept will provide greater assurance that each concept has a well-defined boundary that separates it conceptually from other things. A concept definition should embody criteria for inclusion, as in data modeling (Everest, 1986). This will help ensure that each data instance will be interpreted as an instance of only one construct.

Populate a database of text instances within construct tables (Table 2). Bringing all the instances of a construct together into a table allows the researcher to do a more complete and consistent job of the large volumes of constant comparisons grounded theory requires (concern 2.). Glaser & Strauss (1967) recommend that each instance be compared with other instances and to other constructs as it is categorized. Populating a database facilitates this. Concern 3. is also addressed, in that readers or reviewers could be shown the tables of instances providing evidence for each construct.

Treating each grounded theory construct as though it were an entity with instances, cross-reference (i.e., hyperlink) each tabled construct instance to the text from which the construct was induced. Hyperlinking text to construct will enable one to prevent duplicate coding of text segments (concern 1.). Referencing will also facilitate reliability tests (concern 4.), since a second person can retrace the coding steps and either agree or disagree with them. It will also enable reader or reviewer inspection of the coding interpretations to assess model evidence (concern 3.). Evidence standards can then be applied.

Table 2 Examples of Tables of Data Instances

ACCOUNT (A Data Modeling Entity)

Attributes		
Acct_No	**SSN**	**Balance**
103878	613474974	101.94
187974	999746447	87.86
208445	099874746	375.36

GSS FEATURE (A Grounded Theory Construct)

Attributes (Properties)	
What: Feature Description	**Who Knew the Feature**
to maintain anonymity of comment	Interviewee 16
it documents meetings	Interviewee 27
you can input ideas at the same time	Interviewee 4

Normalize the data by employing a standard similar to third normal form to determine whether the level of validity is adequate (addressing concern 5.). Normalization creates smaller, more unitary relations by analyzing how the attributes relate to each other (Kent, 1983). Applying normalization, one would analyze each grounded theory construct descriptor to assure that it provides a fact about the whole construct and nothing but the construct. For a second normal form test, grounded theorists would examine the glossarized definition of the construct to see if it has one meaning (a unitary definition), or if it has two or more related meanings (a composite definition). If it contains two or more related meanings, the researcher can examine an attribute's instances to see whether they relate to only a part of the composite definition or all of the composite definition. Any violations would indicate less than acceptable construct validity, resulting in rework of the construct. For a third normal form (3NF) test, an attribute's instances would be compared to another attribute's instances within the table to see whether they relate more to the meaning of other attributes than to the meaning of the construct as a whole. The attribute could also be compared with attributes of a similar construct for discriminant validity. Using such techniques, each set of instances can be tested to see if they provide a fact about the construct, the whole construct, and nothing but the construct, raising confidence in the model's validity. This provides a standard for grounded theorists, much as 3NF does for data modelers.

Conclusion and Caveat

We first discussed how grounded theory and data modeling are (imperfectly) analogous methods. Second, we argued that because the two methods are analogous, principles and techniques used in data modeling may improve the reliability and construct validity of grounded theory research, at the same time enabling quality standards.

Three caveats are in order. First, the types of validity-enhancing techniques described here are very time-consuming. When one has large amounts of data, some or all of the techniques should be used on a sample of the data. Second, grounded theory methods may be used to produce action-oriented theories, while the entity-relationship model is static and may also constrain the interpretive richness of grounded theory. Third, the analogy falls apart in these ways: a) first normal form, since it requires unitary data in each table cell, is not applicable to grounded theory; b) as already mentioned, the purposes of grounded theory and data modeling are very different; c) data modeling normalization addresses data anomalies and redundancies, which are not the same as reliability and validity. Still…there's no such thing as a perfect analogy.

Selected References

Brown, R. G. "Data Modeling Methodologies--Contrasts in Style," B. von Halle and D. Kull (Eds.), Data Management Handbook, Auerbach, Boston, 1993, pp. 389-439.

Denzin, N. K. "The art and politics of interpretation." In N. K. Denzin & Y. S. Lincoln (Eds.). Handbook of qualitative research, pp. 500-515. Sage, Newbury Park, CA, 1994.

Everest, G. C. Database Management. McGraw-Hill, New York, 1986.

Ives, K. H. "Advances in Qualitative Methodology." Case Analysis, Volume 1, Number 4, 1981, pp. 357-364.

Kent, W. "A Simple Guide to Five Normal Forms in Relational Database Theory." Communications of the ACM, 26:120-124, 1983.

March, S. T. "Logical Data Modeling." Macmillan Encyclopedia of Computers, 1992.

Schwab, D. P. "Construct validity in organizational behavior." In B. M. Staw & L. L. Cummings (Eds.). Research in Organizational Behavior, volume 2, JAI Press, Greenwich, CN, 1980, pp. 3-43.